TEORIA ERGODICA

Eleonora Catsigeras

Apuntes para el curso de Maestría en Matemática.
Noviembre de 1997.

CAPITULO 7
ENTROPIA.

Sea $(x, \mathcal{A}, \mu)$ un espacio de probabilidad y sea $T : X \mapsto X$ una transformación medible.

Veamos primero una introducción intuitiva de lo que se entiende por cantidad de información.

Sea $A \in \mathcal{A}$ tal que $\mu(A) > 0$. La cantidad de información i de la sentencia "$x \in A$" es, dicho en lenguaje corriente, el espacio que merecería ocupar la noticia "$x \in A$"en un informativo.

Pediremos lo siguiente:

1. La cantidad de información es nula si A ocurre casi siempre, es mayor cuanto menos probable es A y siempre es un número real ≥ 0, que depende solo de la probabilidad $\mu(A)$ de A. Es decir,
$$i : (0,1] \mapsto I\!R^+ \qquad i(p) \quad \text{decreciente con } p \qquad i(1) = 0$$

2. Si dos sucesos A y B son independientes, entonces la cantidad de información de la sentencia conjunta "$x \in A$ y $x \in B$" es igual a la suma de las cantidades de información de "$x \in A$" y de "$x \in B$". Es decir,
$$i(pq) = i(p) + i(q) \qquad \text{para todos } p \text{ y } q \in (0,1]$$

Derivando la condición (2) respecto a p, y multiplicando por p, se obtiene $pqi'(pq) = pi'(p)$, para todos p y q en $(0,1]$. Luego $pi'(p) = -k$ constante para todo p en $(0,1]$. Integrando esta ecuación, y considerando la condición (1) se obtiene

$$i(p) = -k \log p \qquad \text{para } k \text{ constante } > 0$$

Elegir diferentes valores para la constante k, equivale a elegir diferentes unidades para medir la cantidad de información, y equivale también a elegir diferentes bases para el logaritmo. En Teoría Ergódica se utiliza $k = 1$. En Ingeniería se utiliza $k = 1/\log 2$, o lo que es lo mismo, se calculan los logaritmos en base 2. La unidad de medida así obtenida se llama "*bit* ".

Definición 0.1 La cantidad de información de un suceso con probabilidad $p > 0$ es

$$i(p) = -\log p$$

La cantidad de *bits* es $i(p)/\log 2 = -\log_2(p)$.

La palabra *bit* proviene del siguiente ejemplo:

Ejemplo: Sea el shift de Bernoulli de dos símbolos equiprobables $\Sigma(1/2, 1/2)$. Sea A el cilindro que se obtiene al fijar los primeros N símbolos: $A = C(a_1, a_2, \ldots, a_N)$. La probabilidad de A es $\mu(A) = \frac{1}{2^N}$. La cantidad de información de la sentencia "$x \in A$"es

$$i(\frac{1}{2^N}) = -\log(\frac{1}{2^N}) = N \log 2 = N \text{ bits}$$

Sean A y B dos sucesos con probabilidad positiva. La probabilidad condicional de B dado A se define como

$$\mu(B|A) = \frac{\mu(A \cap B)}{\mu(A)}$$

La cantidad de información del suceso B dado A, se define como

$$i(\mu(B|A)) = -\log \mu(B|A) = -\log \mu(A \cap B) + \log \mu(A)$$

Luego

$$i(\mu(A \cap B)) = i(\mu(A)) + i(\mu(B|A))$$

Interpretación: El espacio necesario para trasmitir la información "$x \in A$ y $x \in B$"es la suma del espacio necesario para trasmitir "$x \in A$"más el necesario para trasmitir "$x \in B$ dado que ya se sabe $x \in A$".

1 Entropía de particiones

Definición 1.1 Una *partición* $\mathcal{P}$ de $(X, \mathcal{A}, \mu)$ es una colección de conjuntos $P \in \mathcal{A}$, dos a dos disjuntos, con $\mu(P) > 0$ y $\mu(\bigcup P) = 1$.

Los conjuntos P de $\mathcal{P}$ se llaman *átomos* de la partición. Todos tienen medida positiva y son dos a dos disjuntos. Entonces la partición $\mathcal{P}$ es una colección finita o infinita numerable de átomos.

Se llama *entropía* de la partición $\mathcal{P}$ al promedio ponderado con la medida de cada átomo, de la cantidad de información de los átomos de $\mathcal{P}$. Es decir:

Definición 1.2 *Entropía de la partición* $\mathcal{P}$ es

$$H(\mathcal{P}) = -\sum_{P \in \mathcal{P}} \mu(P) \log \mu(P)$$

Obsérvese que es un número mayor o igual que cero, o infinito. Además $H(\mathcal{P}) = 0$ si y solo si $\mathcal{P}$ tiene un único átomo, luego, con medida uno.

La entropía de la partición es el valor esperado, (es decir la integral respecto a μ), de la función que a casi todo punto $x \in X$ le asigna la cantidad de información del átomo $P_x \in \mathcal{P}$ que contiene a x. Más precisamente:

$$f(x) = -\log \mu(P_x), \quad H(\mathcal{P}) = \int f \, \mathrm{d}\mu$$

Definición 1.3 Dadas dos particiones $\mathcal{P}$ y $\mathcal{Q}$ se define la *partición producto* $\mathcal{P} \bigwedge \mathcal{Q}$ como la formada por todas las intersecciones $P \bigcap Q$, con medida no nula, de $P \in \mathcal{P}$ y $Q \in \mathcal{Q}$.

Notamos que el producto de particiones es una operación conmutativa, asociativa y que tiene como neutro la partición $\{X\}$ (o también cualquier partición con entropía nula).

Se llama entropía condicional de la partición $\mathcal{Q}$ dada la partición $\mathcal{P}$ al promedio, ponderado con la probabilidad $\mu(P \cap Q)$, de la cantidad de información de Q dado P, para $P \in \mathcal{P}$ y $Q \in \mathcal{Q}$, tales que $\mu(P \cap Q) \neq 0$. Es decir:

Definición 1.4 La *entropía condicional* de $\mathcal{Q}$ dada $\mathcal{P}$ es

$$H(\mathcal{Q}|\mathcal{P}) = -\sum_{P \in \mathcal{P}, Q \in \mathcal{Q}, \mu(P \cap Q) \neq 0} \mu(P \cap Q) \log \frac{\mu(P \cap Q)}{\mu(P)}$$

Nótese que es siempre mayor o igual que cero, o infinito.

Para simplificar la notación convenimos en que la letra imprenta indica un átomo de la partición denotada con la misma letra, pero caligráfica. También convenimos en que, si aparece indicado $0 \cdot \log 0$, entonces el término correspondiente en la suma es nulo. Con esta notación se tiene:

$$H(\mathcal{Q}|\mathcal{P}) = -\sum_{P} \sum_{Q} \mu(P \cap Q) \log \frac{\mu(P \cap Q)}{\mu(P)} \tag{1}$$

$$H(\mathcal{Q} \wedge \mathcal{P}) = -\sum_{P} \sum_{Q} \mu(P \cap Q) \log \mu(P \cap Q) \tag{2}$$

Proposición 1.5 *Sean* $\mathcal{P}, \mathcal{Q}$ *y* $\mathcal{R}$ *particiones. Se cumple:*

1. $H(\mathcal{Q}|\mathcal{P}) \leq H(\mathcal{Q})$

2. $H(\mathcal{P} \wedge \mathcal{Q}) = H(\mathcal{P}) + H(\mathcal{Q}|\mathcal{P})$

3. $H(\mathcal{Q}|\mathcal{P} \wedge \mathcal{R}) \leq H(\mathcal{Q}|\mathcal{R})$

Para probar la proposición anterior precisaremos del siguiente lema:

Lema 1.6 (Desigualdad de Jensen) *Sea $\phi : [0,1] \mapsto \mathbb{R}$ una función continua, derivable hasta segundo orden en $(0,1)$ y con derivada segunda negativa.*

Sea λ_i una sucesión de números reales ≥ 0, tales que $\sum_{i=1}^{\infty} \lambda_i = 1$, y sea x_i una sucesión de puntos en $[0,1]$.

Se cumple:

$$\phi\left(\sum_{i=1}^{\infty} \lambda_i x_i\right) \geq \sum_{i=1}^{\infty} \lambda_i \phi(x_i)$$

Prueba: Las series en el enunciado son absolutamente convergentes porque sus términos i-ésimos son menores o iguales, en valor absoluto, que una constante por λ_i, y la serie de los λ_i converge por hipótesis.

Probemos primero que, para todo $n \geq 1$, si $\sum_{i=1}^{n} \rho_i = 1$, con ρ_i dados ≥ 0, entonces se cumple:

$$\phi\left(\sum_{i=1}^{n} \rho_i x_i\right) \geq \sum_{i=1}^{n} \rho_i \phi(x_i) \tag{3}$$

Para $n = 1$ es trivial, porque, en ese caso, $\rho_1 = 1$.

Para $n = 2$: sean x_1 y $x_2 \in [0.1]$. Como la concavidad de la gráfica de la función ϕ es negativa, porque la derivada segunda es negativa, entonces la gráfica entre x_1 y x_2 está por arriba de la cuerda, es decir, del segmento que tiene por extremos los puntos $(x_1, \phi(x_1))$ y $(x_2, \phi(x_2))$. Esto significa que, para todo $0 \leq \rho \leq 1$ se cumple:

$$\phi(\rho x_1 + (1-\rho)x_2) \geq \rho\phi(x_1) + (1-\rho)\phi(x_2) \tag{4}$$

Por lo tanto se sumple la desigualdad (3) cuando $n = 2$.

Supongamos cierta la desigualdad (3) para n. Demostrémosla para $n+1$. Se tiene $\sum_{i=1}^{n+1} \rho_i = 1$. Sea $\rho = \sum_{i=1}^{n} \rho_i$. Entonces se tiene $\rho_{n+1} = 1 - \rho$. Sea $X = \sum_{i=1}^{n} \rho_i x_i / \rho$. Se cumple

$$\sum_{i=1}^{n+1} \rho_i x_i = \rho X + (1-\rho)x_{n+1}$$

Aplicando la desigualdad (4), y luego la hipótesis de inducción, se tiene

$$\phi\left(\sum_{i=1}^{n+1} \rho_i x_i\right) = \phi(\rho X + (1-\rho)x_{n+1}) \geq \rho\phi(X) + (1-\rho)\phi(x_{n+1})$$

$$= \rho\phi\left(\sum_{i=1}^{n}\frac{\rho_i x_i}{\rho}\right) + \rho_{n+1}\phi(x_{n+1}) \geq \rho\sum_{i=1}^{n}\frac{\rho_i\phi(x_i)}{\rho} + \rho_{n+1}\phi(x_{n+1}) = \sum_{i=1}^{n+1}\rho_i\phi(x_i)$$

Por lo tanto vale (3) para todo n natural.

Ahora terminemos la demostración. Como las series son convergentes, son el límite de sus reducidas n-ésimas. Siendo ϕ continua:

$$\phi\left(\sum_{i=1}^{\infty}\lambda_i x_i\right) = \lim_{n\to\infty}\phi\left(\frac{\sum_{i=1}^{n}\lambda_i x_i}{\sum_{i=1}^{n}\lambda_i}\right) \geq \lim\frac{\sum_{i=1}^{n}\lambda_i\phi(x_i)}{\sum_{i=1}^{n}\lambda_i} = \sum_{i=1}^{\infty}\lambda_i\phi(x_i)$$

■

Ahora estamos en condiciones de demostrar la proposición 1.5

Prueba de la proposición 1.5:

La afirmación 1 es consecuencia inmediata de la afirmación 3 tomando $\mathcal{R} = \{X\}$.

Veamos la afirmación 2:

$$H(\mathcal{Q}\wedge\mathcal{P}) = -\sum_P\sum_Q\mu(P\cap Q)\log\mu(P\cap Q)$$

$$= -\sum_P\sum_Q\mu(P\cap Q)\log\frac{\mu(P\cap Q)}{\mu(P)} - \sum_P\sum_Q\mu(P\cap Q)\log\mu(P) =$$

$$= H(\mathcal{Q}|\mathcal{P}) - \sum_P\mu(P)\log\mu(P) = H(\mathcal{Q}|\mathcal{P}) + H(\mathcal{P})$$

Ahora probemos la afirmación 3: Sea

$$\phi_(x) = -x\log x \text{ si } x\in(0,1] \text{ y } \phi(0)=0$$

Es una función continua en [0,1], y tiene derivada segunda negativa en (0,1). Luego, ϕ cumple la desigualdad de Jensen, probada en el lema anterior.

Usando como convenio de notación que son nulos los términos denotados como $0\log 0$, o $0\log(0/0)$, se tiene:

$$H(\mathcal{Q}|\mathcal{P}\wedge\mathcal{R}) = -\sum_R\sum_Q\sum_P\mu(P\cap Q\cap R)\log\frac{\mu(P\cap Q\cap R)}{\mu(P\cap R)}$$

$$= \sum_R\mu(R)\sum_Q\sum_P\frac{\mu(P\cap R)}{\mu(R)}\phi\left(\frac{\mu(P\cap Q\cap R)}{\mu(P\cap R)}\right)$$

Tomemos $Q\in\mathcal{Q}$ y $R\in\mathcal{R}$ fijos. Enumeremos con subíndice $i\geq 1$ todos los conjuntos $P_i\in\mathcal{P}$, tales que $\mu(P_i\cap R)\neq 0$. Definamos los números $\lambda_i = \mu(P_i\cap R)/\mu(R)$ y $x_i = \mu(P_i\cap Q\cap R)/\mu(P_i\cap R)$.

Se cumple que $x_i \in [0,1]$, que $\lambda_i \geq 0$ y que $\sum_{i=1}^{\infty} \lambda_i = 1$. Luego, aplicando la desigualdad de Jensen, se tiene:

$$H(\mathcal{Q}|\mathcal{P} \wedge \mathcal{R}) = \sum_R \mu(R) \sum_Q \sum_{i\geq 1} \lambda_i \phi(x_i) \leq \sum_R \mu(R) \sum_Q \phi\left(\sum_{i\geq 1} \lambda_i x_i\right)$$

Observando que $\sum_{i\geq 1} \lambda_i x_i = \mu(Q \cap R)/\mu(R)$ se concluye:

$$H(\mathcal{Q}|\mathcal{P} \wedge \mathcal{R}) \leq -\sum_R \mu(R) \sum_Q \frac{\mu(Q \cap R)}{\mu(R)} \log\left(\frac{\mu(Q \cap R)}{\mu(R)}\right) = H(\mathcal{Q}|\mathcal{R})$$

■

Corolario 1.7 (Entropía del producto de particiones)

$$H(\mathcal{P}) \leq H(\mathcal{P} \wedge \mathcal{Q}) \leq H(\mathcal{P}) + H(\mathcal{Q})$$

Prueba: Basta recordar que $H(\mathcal{Q}|\mathcal{P}) \geq 0$, y usar la igualdad 2) y la desigualdad 1) de la proposición 1.5. ■

Corolario 1.8 (Entropía del producto condicionada) *Si* $H(\mathcal{R}) < \infty$ *entonces:*

i) $H(\mathcal{P} \wedge \mathcal{Q}|\mathcal{R}) = H(\mathcal{P}|\mathcal{R}) + H(\mathcal{Q}|\mathcal{P} \wedge \mathcal{R})$

ii) $H(\mathcal{P}|\mathcal{R}) \leq H(\mathcal{P} \wedge \mathcal{Q}|\mathcal{R}) \leq H(\mathcal{P}|\mathcal{R}) + H(\mathcal{Q}|\mathcal{R})$

Prueba: i) Aplicando tres veces la igualdad (2) de 1.5:

$H(\mathcal{R} \wedge (\mathcal{P} \wedge \mathcal{Q})) = H(\mathcal{R}) + H(\mathcal{P} \wedge \mathcal{Q}|\mathcal{R})$

$H((\mathcal{P} \wedge \mathcal{R}) \wedge \mathcal{Q}) = H(\mathcal{P} \wedge \mathcal{R}) + H(\mathcal{Q}|\mathcal{P} \wedge \mathcal{R}) = H(\mathcal{R}) + H(\mathcal{P}|\mathcal{R}) + H(\mathcal{Q}|\mathcal{P} \wedge \mathcal{R})$

Como $\mathcal{R} \wedge \mathcal{P} \wedge \mathcal{Q} = \mathcal{P} \wedge \mathcal{R} \wedge \mathcal{Q}$, igualando las expresiones de los últimos miembros, se tiene el resultado enunciado.

ii) De i), siendo siempre $H(\mathcal{Q}|\mathcal{P} \wedge \mathcal{R}) \geq 0$, se tiene $H(\mathcal{P} \wedge \mathcal{Q}|\mathcal{R}) \geq H(\mathcal{P}|\mathcal{R})$.

Aplicando la desigualdad (3) de la proposición 1.5, a (i), se tiene $H(\mathcal{P} \wedge \mathcal{Q}|\mathcal{R}) \leq H(\mathcal{P}|\mathcal{R}) + H(\mathcal{Q}|\mathcal{R})$. ■

Nota: Las partes i) y ii) del corolario 1.8 son también ciertas aunque $H(\mathcal{R}) = \infty$, pero para probarlas hay que remitirse a las definiciones.

Corolario 1.9 (Diferencia de entropías) *Si* $H(\mathcal{P}) < \infty$ *, entonces, para cualquier otra partición* $\mathcal{Q}$ *se cumple*

$$H(\mathcal{Q}) - H(\mathcal{P}) \leq H(\mathcal{Q}|\mathcal{P})$$

Prueba: Por el corolario 1.7 y la igualdad (2) de la proposición 1.5, se tiene:

$$H(\mathcal{Q}) \leq H(\mathcal{P} \wedge \mathcal{Q}) = H(\mathcal{P}) + H(\mathcal{Q}|\mathcal{P})$$

Restando $H(\mathcal{P})$ se obtiene la tesis. ■

2 Refinamiento de particiones

Definición 2.1 Sean $\mathcal{P}$ y $\mathcal{Q}$ dos particiones. Se dice que $\mathcal{Q}$ *es más fina que* $\mathcal{P}$, o $\mathcal{P}$ *más gruesa que* $\mathcal{Q}$, si todo átomo de $\mathcal{P}$ coincide μ-c.t.p. con una unión de átomos de $\mathcal{Q}$.

El producto de dos particiones es, obviamente, más fino que cada una de las particiones.

Dos particiones *coinciden* μ- c.t.p. si cada átomo de una de ellas es igual μ-c.t.p. a algún átomo de la otra.

Nótese que $\mathcal{Q}$ es más fina que $\mathcal{P}$ si y solo si $\mathcal{P} \wedge \mathcal{Q}$ coincide μ-c.t.p. con $\mathcal{Q}$. Como el valor de la entropía depende solo de la medida de los átomos, en todo cálculo de entropías, donde aparezca la partición $\mathcal{Q}$, podemos poner $\mathcal{P} \wedge \mathcal{Q}$.

Cuando $\mathcal{Q}$ es más fina que $\mathcal{P}$, se cumple $H(\mathcal{Q} \wedge \mathcal{P}) = H(\mathcal{Q})$. Luego, aplicando la propiedad 2) de la proposición 1.5,

$$H(\mathcal{Q}) = H(\mathcal{Q} \wedge \mathcal{P}) = H(\mathcal{Q}) + H(\mathcal{P}|\mathcal{Q})$$

se concluye que si $\mathcal{Q}$ tiene entropía finita y es más fina que $\mathcal{P}$, entonces $H(\mathcal{P}|\mathcal{Q}) = 0$. (Aplicando la definición directamente, se puede demostrar que el resultado también vale aunque $H(\mathcal{Q}) = \infty$).

Veamos ahora que la entropía crece cuanto más fina es la partición. Si se condiciona, la entropía crece cuanto mas fina es la partición condicionada, y decrece cuanto más fina es la partición condicionante:

Proposición 2.2 *Sean $\mathcal{P}, \mathcal{Q}, \mathcal{R}$ tres particiones. Si la partición $\mathcal{Q}$ es más fina que $\mathcal{P}$, entonces:*

1. $H(\mathcal{Q}) \geq H(\mathcal{P})$
2. $H(\mathcal{Q}|\mathcal{R}) \geq H(\mathcal{P}|\mathcal{R})$
3. $H(\mathcal{R}|\mathcal{Q}) \leq H(\mathcal{R}|\mathcal{P})$

Prueba: Para demostrar 1) apliquemos el corolario 1.7:

$$H(\mathcal{P}) \leq H(\mathcal{P} \wedge \mathcal{Q}) = H(\mathcal{Q})$$

Para demostrar 2) usemos la desigualdad ii) del corolario 1.8:

$$H(\mathcal{Q}|\mathcal{R}) = H(\mathcal{Q} \wedge \mathcal{P}|\mathcal{R}) \geq H(\mathcal{P}|\mathcal{R})$$

Para demostrar 3) usemos la parte 3) de la proposición 1.5:

$$H(\mathcal{R}|\mathcal{Q}) = H(\mathcal{R}|\mathcal{P} \wedge \mathcal{Q}) \leq H(\mathcal{R}|\mathcal{P})$$

■

Definición 2.3 Sea $\{\mathcal{P}_n\}_{n\geq 1}$ una sucesión de particiones. Indicamos con

$$\bigvee_{n\geq 1} \mathcal{P}_n$$

a la mínima σ-álgebra que contiene a todos los átomos de todas las particiones de la sucesión. Es decir contiene a $\bigcup_{n\geq 1} \mathcal{P}_n$.

Por lo tanto, una partición $\mathcal{P}$ está contenida en $\bigvee_{n\geq 1} \mathcal{P}_n$, si todos sus átomos pertenecen a esa σ-álgebra.

Decimos que una partición *está contenida* μ-c.t.p. en $\bigvee_{n\geq 1} \mathcal{P}_n$, si todos sus átomos son iguales μ-c.t.p. a algún subconjunto miembro de esa σ-álgebra.

Ahora probemos un lema para demostrar el teorema de Kolmogorov-Sinai que veremos en las próximas secciones:

Lema 2.4 *Sea $\{\mathcal{P}_n\}_{n\geq 1}$ una sucesión de particiones tales que, para todo $n \geq 1$ la partición $\mathcal{P}_{n+1}$ es mas fina que $\mathcal{P}_n$.*

Sea $\mathcal{P}$ una partición finita contenida μ-c.t.p. en $\bigvee_{n\geq 1} \mathcal{P}_n$.

Entonces

$$\lim_{n\to\infty} H(\mathcal{P}|\mathcal{P}_n) = 0$$

Prueba: Todas las particiones consideradas lo son del espacio de probabilidad $(X, \mathcal{A}, \mu)$.

Definamos la familia $\mathcal{F} = \{A \in \mathcal{A}:$ Para todo $\epsilon > 0$ existen $N \geq 1$ y alguna unión B de átomos de $\mathcal{P}_N$ que cumple $\mu(A\Delta B) < \epsilon\}$.

Nótese que, si $A \in \mathcal{F}$, entonces, para todo $n \geq N$, existe alguna unión $B^{(n)}$ de átomos de $\mathcal{P}_n$ que cumple $\mu(A\Delta B^{(n)}) < \epsilon$. Esto se debe a que la partición $\mathcal{P}_n$ es mas fina que $\mathcal{P}_N$. Por lo tanto, si N es adecuado, cualquier natural mayor que N también lo es.

Es inmediato ver que la familia $\mathcal{F}$ contiene a $\mathcal{P}_N$, para todo $N \geq 1$.

Afirmamos que $\mathcal{F}$ es una σ-álgebra. En efecto, para probar que el complemento de A está en $\mathcal{F}$, basta usar los mismos N y B que para A, y comparar A^c con la unión de los átomos de la partición $\mathcal{P}_N$ que no están en B.

Ahora, probemos que la unión numerable de conjuntos $A_i \in \mathcal{F}$ también pertence a $\mathcal{F}$:

Sea $\epsilon > 0$ dado. Para cada $i \geq 1$ sean $N_i \geq 1$ y B_i, unión de átomos de $\mathcal{P}_{N_i}$, que cumple $\mu(A_i \Delta B_i) < \epsilon/2^{i+1}$ Sea $B = \bigcup_{i\geq 1} B_i$. Se tiene:

$$\mu(A\Delta B) \leq \sum_{i=1}^{\infty} \mu(A_i \Delta B_i) < \sum_{i=1}^{\infty} \epsilon/2^{i+1} = \epsilon/2 \tag{5}$$

Para todo $k \geq 1$ defínase $C_k = \bigcup_{i=1}^{k} B_i$. Se cumple $C_1 \subset C_2 \subset \ldots \subset C_k \subset \ldots \bigcup_{i\geq 1} C_i = B$. Luego $\mu(B) = \lim \mu(C_k)$, y existe $K \geq 1$ tal que $\mu(B) - \mu(C_K) < \epsilon/2$. Como $C_K \subset B$, se tiene:

$$\mu(B\Delta C_K) < \epsilon/2 \tag{6}$$

Sea $N = \max_{1\leq i\leq K} N_i$. Como B_i es una unión de átomos de $\mathcal{P}_{N_i}$, coincide μ-c.t.p. con una unión de átomos de $\mathcal{P}_N$. Luego $C_K = \bigcup_{i=1}^{K} B_i$ también coincide μ-c.t.p. con una unión H de átomos de $\mathcal{P}_N$.

Por (5) y (6):

$$\mu(A\Delta H) = \mu(A\Delta C_K) \leq \mu(A\Delta B) + \mu(B\Delta C_k) < \epsilon$$

Luego $A \in \mathcal{F}$ y $\mathcal{F}$ es una σ-álgebra.

Por lo tanto la familia $\mathcal{F}$ contiene a la σ-álgebra $\bigvee_{n\geq 1} \mathcal{P}_n$. Esta, por hipótesis, contiene μ-c.t.p. a la partición $\mathcal{P}$. Si $A \in \mathcal{F}$, entonces cualquier otro conjunto que coincida μ- c.t.p. con A también pertence a $\mathcal{F}$. Luego, todos los átomos de $\mathcal{P}$ pertenecen a la familia $\mathcal{F}$.

Siendo $\mathcal{P}$ una partición finita, se tiene $\mathcal{P} = \{P_1, P_2, \ldots P_r\}$, con $P_i \in \mathcal{F}$, para todo $i = 1, 2, \ldots, r$. Dado ϵ existe $N = N(\epsilon) \geq 1$ (el mismo para todo $i = 1, 2, \ldots, r$) y existe, para cada $i = 1, 2, \ldots, , r$, alguna unión Q_i de átomos de $\mathcal{P}_N$, tal que

$$\mu(P_i \Delta Q_i) < \epsilon$$

Los conjuntos Q_i no forman necesariamente una partición, porque pueden intersectarse, aunque los P_i no lo hacen. Pero aproximan a los átomos de la partición $\mathcal{P}$ y son uniones de átomos de $\mathcal{P}_N$.

Construyamos, usando los Q_i, una partición $\mathcal{S} = \mathcal{S}(\epsilon)$ cuyos átomos continúen aproximando a los P_i y sigan siendo μ-c.t.p. uniones de átomos de $\mathcal{P}_N$. Entonces $\mathcal{S}$ será mas gruesa que $\mathcal{P}_N$, y por la proposición 2.2 se cumple:

$$0 \leq H(\mathcal{P}|\mathcal{P}_n) \leq H(\mathcal{P}|\mathcal{P}_N) \leq H(\mathcal{P}|\mathcal{S}(\epsilon)) \quad \forall\, n \geq N = N(\epsilon)$$

Para demostrar la tesis $H(\mathcal{P}|\mathcal{P}_N) \to 0$, bastará entonces probar que $H(\mathcal{P}|\mathcal{S}(\epsilon))_{\epsilon\to\infty} \to 0$

Pasemos a la construcción de $\mathcal{S}(\epsilon)$, definiendo

$$S_1 = Q_1, \quad S_i = Q_i - \bigcup_{j=1}^{i-1} Q_j, \quad \text{para } 2 \leq i \leq r-1, \quad S_r = \left(\bigcup_{j=1}^{r-1} Q_j\right)^c$$

$$\mathcal{S}(\epsilon) = \{S_i : 1 \leq i \leq r, \quad \mu(S_i) \neq 0\}$$

Por construcción $\mathcal{S}(\epsilon)$ es una partición, y sus átomos son μ-c.t.p. uniones de átomos de $\mathcal{P}_N$, porque los Q_i lo eran.

Afirmamos que para todo $i = 1, 2, \ldots, r$ se cumple

$$\mu(P_i \Delta S_i) < r\epsilon$$

En efecto, para $i = 1 : S_1 = Q_1$ y $\mu(P_1 \Delta Q_1) < \epsilon$.

Para $2 \leq i \leq r-1$, si $1 \leq j \leq i-1$ se tiene:

$$S_i \subset Q_i, \quad S_i^c = Q_i^c \bigcup \bigcup_{j=1}^{i-1} Q_j, \quad P_i \subset P_j^c$$

Luego $\mu(P_i \Delta S_i) = \mu(P_i^c \cap S_i) + \mu(P_i \cap S_i^c) \leq \mu(P_i^c \cap Q_i) + \mu(P_i \cap Q_i^c) + \sum_{j=1}^{i-1} \mu(P_j^c \cap Q_j) \leq \sum_{j=1}^{i} \mu(P_j \Delta Q_j) < r\epsilon$.

Finalmente, para $i = r$, si $1 \leq j \leq r-1$, se tiene

$$S_r \subset Q_j^c, \quad S_r^c = \bigcup_{j=1}^{r-1} Q_j, \quad P_i \subset P_j^c, \quad P_r^c = \bigcup_{j=1}^{r-1} P_j \ \mu\text{-c.t.p.}$$

Luego $\mu(P_r \Delta S_r) = \mu(P_r^c \cap S_r) + \mu(P_r \cap S_r^c) \leq \sum_{j=1}^{r-1} \mu(P_j \cap Q_j^c) + \sum_{j=1}^{r-1} \mu(P_j^c \cap Q_j) \leq \sum_{j=1}^{r-1} \mu(P_j \Delta Q_j) < r\epsilon$. Con esto acabamos de demostrar que los átomos de la partición $\mathcal{S}(\epsilon)$ aproximan a los átomos de la partición dada $\mathcal{P}$.

Para terminar, falta probar que $H(\mathcal{P}|\mathcal{S}(\epsilon)) \to 0$ cuando $\epsilon \to 0$.

$$H(\mathcal{P}|\mathcal{S}(\epsilon)) = -\sum_i \sum_j \mu(P_j \cap S_i) \log\left(\frac{\mu(P_j \cap S_i)}{\mu(S_i)}\right) = \sum_i \mu(S_i) \sum_j \phi\left(\frac{\mu(P_j \cap S_i)}{\mu(S_i)}\right)$$

donde $\phi(x) = -x \log x$, la primera suma es para aquellos $i \in \{1, 2, \ldots, r\}$ tales que $\mu(S_i) \neq 0$, y la segunda suma para aquellos $j \in \{1, 2, \ldots, r\}$ tales que $\mu(P_j \cap S_i) \neq 0$.

Como hay una cantidad finita acotada de sumandos, basta probar que cada uno converge a cero cuando $\epsilon \to 0$.

Primero veamos que $\mu(S_i) \to \mu(P_i) > 0$. En efecto, $|\mu(S_i) - \mu(P_i)| \leq \mu(S_i \Delta P_i) < r\epsilon \to 0$.

Por un lado, si $j = i$: $\mu(P_i \cap S_i) = \mu(S_i) - \mu(P_i^c \cap S_i)$,

$$0 \leq \mu(P_i^c \cap S_i) \leq \mu(P_i \Delta S_i) < r\epsilon \to 0$$

Entonces

$$\lim_{\epsilon \to 0} \phi\left(\frac{\mu(P_i \cap S_i)}{\mu(S_i)}\right) = \phi(1) = 0$$

Por otro lado, si $j \neq i$, entonces $P_j \subset P_i^c$, y

$$0 \leq \mu(P_j \cap S_i) \leq \mu(P_i^c \cap S_i) \leq \mu(P_i \Delta S_i) < r\epsilon \to 0$$

Luego,

$$\lim_{\epsilon \to 0} \phi\left(\frac{\mu(P_j \cap S_i)}{\mu(S_i)}\right) = \lim_{x \to 0} \phi(x) = 0$$

■

3 Entropía de una transformación respecto a una partición

Sea $T : X \mapsto X$ medible que preserva una medida de probabilidad μ. Sea $\mathcal{P}$ una partición de X con entropía $H(\mathcal{P})$ finita. Es fácil ver que la familia formada por las preimágenes por T de los átomos de $\mathcal{P}$ forman una nueva partición, que denotamos como $T^{-1}\mathcal{P}$, que tiene la misma entropía que $\mathcal{P}$.

$$H(T^{-1}\mathcal{P}) = H(\mathcal{P})$$

Si se tienen dos particiones $\mathcal{P}$ y $\mathcal{Q}$, se cumple:

$$H(T^{-1}\mathcal{P}|T^{-1}\mathcal{Q}) = H(\mathcal{P}|\mathcal{Q})$$

Considérese para cada $n \geq 1$, la partición producto:

$$\mathcal{P}_n = \mathcal{P} \wedge T^{-1}\mathcal{P} \wedge \ldots \wedge T^{-(n-1)}\mathcal{P}$$

Al aumentar n se tienen particiones cada vez mas finas, luego $H(\mathcal{P}_n)$ es creciente con n.

Además, por el corolario 1.7

$$H(\mathcal{P}_n) \leq H(\mathcal{P}) + H(T^{-1}\mathcal{P}) + \ldots + H(T^{-(n-1)}\mathcal{P}) = nH(\mathcal{P})$$

Luego:

$$\frac{H(\mathcal{P}_n)}{n} \leq H(\mathcal{P}) < \infty \quad \forall n \geq 1$$

Proposición 3.1 *La sucesión* $\{H(\mathcal{P}_n)\}_{n\geq 1}$ *es subaditiva, esto es:*

$$H(\mathcal{P}_{n+m}) \leq H(\mathcal{P}_n) + H(\mathcal{P}_m) \quad \forall\, n, m \geq 1$$

Prueba:

$$H(\mathcal{P}_{n+m}) = H\left((\wedge_{j=0}^{n-1} T^{-j}\mathcal{P}) \wedge (\wedge_{j=n}^{n+m-1} T^{-j}\mathcal{P})\right)$$

$$\leq H(\mathcal{P}_n) + H(T^{-n}\mathcal{P}_m) = H(\mathcal{P}_n) + H(\mathcal{P}_m) \quad \forall\, n, m \geq 1$$

■

Teorema 3.2

$$\lim_{n\to\infty} \frac{H(\mathcal{P}_n)}{n} = \inf_{n\geq 1} \frac{H(\mathcal{P}_n)}{n} \leq H(\mathcal{P})$$

Prueba: Basta demostrar que toda sucesión $\{a_n\}_{n\geq 1}$ de reales, subaditiva (es decir $a_{n+m} \leq a_n + a_m$ para todos n y $m \geq 1$), con a_n/n acotada inferiormente, cumple

$$\lim_{n\to\infty} \frac{a_n}{n} = \inf_{n\geq 1} \frac{a_n}{n}$$

Sea $c = inf_{n\geq 1} a_n/n$. Sea, dado $\epsilon > 0$, un natural $n_0 \geq 1$ tal que $a_{n_0}/n_0 < c + \epsilon$.

Si $n \geq n_0$, haciendo la división entera $n = qn_0 + r$, con $0 \leq r < n_0$, y aplicando la subaditividad, se obtiene:

$$\frac{a_n}{n} = \frac{a_{qn_0+r}}{n} \leq \frac{qa_{n_0}}{qn_0 + r} + \frac{a_r}{n} \leq \frac{q}{qn_0} a_{n_0} + \frac{\max_{0\leq i<n_0} a_i}{n} < c + 2\epsilon \quad \forall\, n \geq N$$

si N es suficientemente grande.

Luego

$$c \leq \frac{a_n}{n} < c + 2\epsilon \quad \forall\, n \geq N$$

Es decir

$$\frac{a_n}{n}\,_{n\to\infty} \to c$$

■

Definición 3.3 Si T es medible y $\mathcal{P}$ es una partición con entropía finita, se llama *entropía de la transformación* T *respecto a la partición* $\mathcal{P}$ al número real

$$h(T, \mathcal{P}) = \lim_{n\to\infty} \frac{H(\mathcal{P}_n)}{n} = \inf_{n\geq 1} \frac{H(\mathcal{P}_n)}{n}$$

Interpretación:

$T : X \mapsto X$ define un *sistema dinámico determinístico:* Da una ley conocida, que en función de cada *estado inicial* $x \in X$, determina la sucesión de estados futuros: $T(x), T^2(x), \ldots, T^j(x), \ldots$ (la órbita por x). Supongamos que la transformación T se aplica a cada segundo : $T^j(x)$ es el estado al comienzo del segundo j-ésimo.

El físico o el ingeniero quiere informar sobre el estado del sistema en cada segundo futuro, a partir de su estado inicial x, por lo menos para casi todo punto x. Modestamente, no requiere determinar exactamente el punto $T^j(x)$ del espacio, sino que dispone de una cierta partición $\mathcal{P}$ del espacio, predeterminada, y solo requiere saber en cuál átomo A_j de la partición está el punto $T^j(x)$, para cada instante futuro $j \geq 0$. (Por ejemplo si X es un subconjunto de $\mathbb{R}^2$, quiere tener el dato de qué cuadrante contiene a $T^j(x)$ para cada $j \geq 0$.) La sucesión de átomos A_j, se llama *itinerario* de x. También modestamente, puede tolerar atrasos muy grandes en trasmitir la información, con tal de que, cada tanto, puedan darse los datos acumulados.

El valor esperado de la cantidad de información inicial es $H(\mathcal{P})$, según la interpretación que dimos al definir entropía de una partición, en 1.1.

Al cabo de $n-1$ segundos tendrá la información acumulada:

$$x \in A_0, T(x) \in A_1, T^2(x) \in A_2, \ldots, T^{n-1}(x) \in A_{n-1}$$

Esto es lo mismo que el dato de qué átomo de la partición $\mathcal{P}_n$ contiene al punto x. Luego, el valor esperado de la cantidad de información acumulada al cabo de $n-1$ segundos, será $H(\mathcal{P}_n)$.

Tuvo n segundos para trasmitir esa información acumulada. Entonces debió hacerlo con una *velocidad* promedio $H(\mathcal{P}_n)/n$. Como puede tolerar atrasos tan grandes como necesite, puede acercarse tanto al límite de $H(\mathcal{P}_n)/n$ como desee. Este límite es un ínfimo, y es, por definición la entropía de T respecto a la partición $\mathcal{P}$.

Luego, *la entropía de T respecto a la partición $\mathcal{P}$, es el límite inferior teórico de la velocidad de transmisión de datos requerida para informar sobre el itinerario del sistema.*

$$\text{Cantidad de bits/segundo } = h(t, \mathcal{P})/\log 2$$

Si la entropía es positiva, quiere decir que se requiere de una velocidad promedio positiva, para ir trasmitiendo los datos, a medida que se acumulan. Usando mayor velocidad pueden adelantarse algunos datos. Pero no se pueden trasmitir en tiempo acotado los datos para *todo* el futuro. Si la entropía es positiva, entonces no se pueden hacer previsiones en el estado asintótico futuro, trasmitiendo solo una cantidad finita de datos en tiempo acotado N (porque, si así fuera, la velocidad promedio tendería a cero cuando n tiende a infinito, $n \geq N$).

En este sentido, cuando tiene entropía positiva el sistema es *imprevisible,* aunque sea determinístico.

También podemos dar la siguiente interpretación: si la entropía $h = h(T, \mathcal{P})$ es positiva el sistema es *caótico,* esto es, sensible a las condiciones iniciales, aunque en un sentido débil: hay itinerarios asintóticamente diferentes. En efecto, si el itinerario a partir de cierto instante N no dependiera más del estado inicial x, entonces las particiones $\mathcal{P}_n$, para $n \geq N$ coincidirían con $\mathcal{P}_N$. Al dividir su entropía $H(\mathcal{P}_n)$ entre n, y hacer n tender a infinito, la entropía h de la transformación sería nula.

También podemos dar la siguiente interpretación a la entropía $h = h(T, \mathcal{P})$ de la transformación: Es el grado de *desorden* que provoca la transformación T en el espacio. En efecto, nótese que $H(\mathcal{P}_n)$ es creciente con n, y h es la velocidad con la que crece en relación a n. Si h es positivo, entonces la partición inicial $\mathcal{P}$ debe sufrir más y más subatomizaciones al aplicarle T sucesivas veces (pues de lo contrario, las particiones $\mathcal{P}_n$ serían todas iguales a $\mathcal{P}_N$ para $n \geq N$ y luego h sería cero).

Proposición 3.4 *Cualesquiera sean las particiones $\mathcal{P}$ y $\mathcal{Q}$ con entropías finitas, se cumple:*

1. $h(T, \mathcal{Q}) - h(T, \mathcal{P}) \leq H(\mathcal{Q}|\mathcal{P})$

2. *Si $\mathcal{Q}$ es mas fina que $\mathcal{P}$ entonces* $0 \leq h(T, \mathcal{Q}) - h(T, \mathcal{P})$.

Prueba: 1) Según el corolario 1.9 que acota la diferencia de entropías de particiones, se tiene

$$H(\wedge_{j=0}^{n-1} T^{-j}\mathcal{Q}) - H(\wedge_{j=0}^{n-1} T^{-j}\mathcal{P}) \leq H(\wedge_{j=0}^{n-1} T^{-j}\mathcal{Q})| \wedge_{j=0}^{n-1} T^{-j}\mathcal{P})$$

Aplicando la parte 2) del corolario 1.8, sobre la entropía del producto condicionada:

$$H(\wedge_{j=0}^{n-1} T^{-j}\mathcal{Q})| \wedge_{j=0}^{n-1} T^{-j}\mathcal{P}) \leq \sum_{j=0}^{n-1} H(T^{-j}\mathcal{Q}| \wedge_{i=0}^{n-1} T^{-i}\mathcal{P})$$

Como al condicionar con una partición mas fina disminuye la entropía, según la afirmación 3) de la proposición 2.2:

$$\sum_{j=0}^{n-1} H(T^{-j}\mathcal{Q}| \wedge_{i=0}^{n-1} T^{-i}\mathcal{P}) \leq \sum_{j=0}^{n-1} H(T^{-j}\mathcal{Q}|T^{-j}\mathcal{P}) = \sum_{j=0}^{n-1} H(\mathcal{Q}|\mathcal{P}) = nH(\mathcal{Q}|\mathcal{P})$$

Luego:

$$\frac{1}{n} H(\wedge_{j=0}^{n-1} T^{-j}\mathcal{Q}) - \frac{1}{n} H(\wedge_{j=0}^{n-1} T^{-j}\mathcal{P}) \leq H(\mathcal{Q}|\mathcal{P})$$

y haciendo n tender a infinito, se tiene

$$h(T, \mathcal{Q}) - h(T, \mathcal{P}) \leq H(\mathcal{Q}|\mathcal{P})$$

2) Si $\mathcal{Q}$ es mas fina que $\mathcal{P}$, entonces $\wedge_{j=0}^{n-1} T^{-j}\mathcal{Q}$ es mas fina que $\wedge_{j=0}^{n-1} T^{-j}\mathcal{P}$, y entonces por la proposición 2.2

$$0 \leq H(\wedge_{j=0}^{n-1} T^{-j}\mathcal{Q}) - H(\wedge_{j=0}^{n-1} T^{-j}\mathcal{P})$$

Dividiendo entre n y haciendo n tender a infinito se tiene: $0 \leq h(T, \mathcal{Q}) - h(T, \mathcal{P})$. ■

Proposición 3.5 *1.* $h(T, \mathcal{P}) = h(T, T^{-1}\mathcal{P}) = h(T, T^{-n}\mathcal{P})$ *para todo* $n \geq 0$

2. $h(T, \mathcal{P}) = h(T, \wedge_{j=0}^{m} T^{-j}\mathcal{P})$ *para todo* $m \geq 0$.

3. Si T *es invertible con inversa medible:* $h(T, \mathcal{P}) = h(T, T\mathcal{P}) = h(T, T^{n}\mathcal{P})$ *para todo* $n \in \mathbb{Z}$.

Prueba:

1)$H(\wedge_{j=0}^{n-1} T^{-j}\mathcal{P}) = H(T^{-1} \wedge_{j=0}^{n-1} T^{-j}\mathcal{P}) = H(\wedge_{j=0}^{n-1} T^{-j}(T^{-1}\mathcal{P}))$

Dividiendo entre n y haciendo n tender a infinito se tiene $h(T, \mathcal{P}) = h(T, T^{-1}\mathcal{P})$. La otra igualdad se obtiene aplicando n veces la primera.

2) $\wedge_{i=0}^{n-1} T^{-i}(\wedge_{j=0}^{m} T^{-j}\mathcal{P}) = \wedge_{i=0}^{n+m-1} T^{-i}\mathcal{P}$.

Calculando su entropía H, dividiendo entre n y haciendo n tender a infinito, con m fijo, se tiene la igualdad 2).

3) Si T es invertible y preserva μ, entonces T^{-1} también preserva μ. Luego, para toda partición $\mathcal{P}$ se cumple $H(T\mathcal{P}) = H(\mathcal{P})$. La prueba sigue análogamente a la de 1), tomando la imagen por T de $\mathcal{P}$ en lugar de su preimagen. ■

4 Entropía métrica de una transformación

Definición 4.1 Sea $T : X \mapsto X$ medible que preserva la medida de probabilidad μ. Se llama *entropía métrica de* T *para la medida* μ a

$$h_{\mu}(T) = \sup\{h(T, \mathcal{P}) : \mathcal{P} \text{ partición tal que } H(\mathcal{P}) < \infty\} \in [0, \infty]$$

Teorema 4.2

$$h_{\mu}(T) = \sup\{h(T, \mathcal{P}) : \mathcal{P} \text{ partición finita }\}$$

Prueba: Como toda partición finita tiene entropía finita, se cumple

$$h_\mu(T) \geq \sup\{h(T,\mathcal{P}) : \mathcal{P} \text{ partición finita }\}$$

Para probar la desigualdad contraria, basta demostrar que dada una partición $\mathcal{P}$ con $H(\mathcal{P}) < \infty$, existe una sucesión $\mathcal{P}_n$ de particiones finitas tales que $h(T,\mathcal{P}_n) \to h(T,\mathcal{P})$.

Dada $\mathcal{P} = \{P_i\}_{i\geq 1}$ sea $\mathcal{P}_n = \{P_1,\ldots,P_{n-1},\bigcup_{i\geq n} P_i\}$.

$\mathcal{P}$ es mas fina que $\mathcal{P}_n$. Luego vale la proposición 3.4:

$$0 \leq h(T,\mathcal{P}) - h(T,\mathcal{P}_n) \leq H(\mathcal{P}|\mathcal{P}_n)$$

y basta probar que $H(\mathcal{P}|\mathcal{P}_n) \to 0$.

Aplicando la definición de entropía condicionada, y observando que para $i = 1,2,\ldots,n-1$, los átomos P_i de las particiones $\mathcal{P}$ y $\mathcal{P}_n$ coinciden, se tiene:

$$H(\mathcal{P}|\mathcal{P}_n) = -\sum_{i\geq n} \mu(P_i)\log\frac{\mu(P_i)}{\mu(\bigcup_{j\geq 1} P_j)} = \sum_{i\geq n}\phi(\mu(P_i)) - \phi\left(\sum_{j\geq n}\mu(P_j)\right)$$

donde $\phi(x) = -x\log x$.

Como $H(\mathcal{P}) = \sum_{i\geq 1}\phi(\mu(P_i)) < \infty$, la cola de la serie $\sum_{i\geq n}\phi(\mu(P_i))$ converge a cero cuando n tiende a infinito.

Por otro lado, como $\sum_{j\geq 1}\mu(P_j) = 1$, la cola de la serie $\sum_{j\geq n}\mu(P_j)$ converge a cero cuando n tiende a infinito. Siendo $\lim_{x\to 0}\phi(x) = 0$, se tiene que $\phi\left(\sum_{j\geq n}\mu(P_j)\right)$ converge a cero cuando n tiende a infinito. ■

Teorema 4.3 *Sean $T_1 : X_1 \mapsto X_1$ que preserva la medida μ_1, y $T_2 : X_2 \mapsto X_2$ que preserva la medida μ_2.*

Si T_1 y T_2 son equivalentes por isomorfismo de espacios de medida, entonces las entropías métricas son iguales:

$$h_{\mu_1}(T_1) = h_{\mu_2}(T_2)$$

Prueba: Sea $f : X_1 \mapsto X_2$ el isomorfismo de espacios de medida tal que $f\circ T_1 = T_2 \circ f$. Para toda partición $\mathcal{P}_2 = \{P_i\}_{i\geq 1}$ de X_2, la partición $f^{-1}\mathcal{P}_2 = \{f^{-1}(P_i)\}_{i\geq 1}$ de X_1, cumple $H(f^{-1}\mathcal{P}_2) = H(\mathcal{P}_2)$, porque f preserva medidas.

Además, como $f\circ T_1^j = T_2^j \circ f$ para todo $j \geq 0$, se cumple

$$f^{-1}\wedge_{j=0}^{n-1} T_2^{-j}\mathcal{P}_2 = \wedge_{j=0}^{n-1} T_1^{-j} f^{-1}\mathcal{P}_2$$

de donde se obtiene que

$$H(\wedge_{j=0}^{n-1} T_2^{-j}\mathcal{P}_2) = H(\wedge_{j=0}^{n-1} T_1^{-j} f^{-1}\mathcal{P}_2)$$

Dividiendo entre n y haciendo n tender a infinito se obtiene $h(T_2, \mathcal{P}_2) = h(T_1, f^{-1}\mathcal{P}_2)$.

Como vale para cualquier partición $\mathcal{P}_2$ de X_2, se tiene:

$$h_{\mu_2}(T_2) = \sup\{h(T_2, \mathcal{P}_2) : H(\mathcal{P}_2) < \infty\} \leq \sup\{h(T_1, \mathcal{P}_1) : H(\mathcal{P}_1) < \infty\} = h_{\mu_1}(T_1)$$

Como la relación de ser equivalentes por isomorfismo de espacios de medida es simétrica, usando f^{-1} en lugar de f, se obtiene la desigualdad contraria. ■

La entropía métrica de una transformación es un supremo de las entropías de la transformación respecto a particiones. Para calcularla puede ser muy útil saber cuando es un máximo, y para que partición se alcanza ese máximo. Condiciones que aseguran eso estarán dadas en el teorema de Kolmogorov-Sinai, para el cual necesitamos la siguiente definición:

Definición 4.4 Sea T medible en un espacio de medida $(X, \mathcal{A}, \mu)$. Una partición $\mathcal{P}$ se llama *generadora* si se cumple alguna de las siguientes condiciones:

i) o bien $\bigvee_{j=0}^{\infty} T^{-j}\mathcal{P} = \mathcal{A}$ μ-c.t.p.

ii) o bien $\bigvee_{j=0}^{\infty} T^{-j}\mathcal{P} \neq \mathcal{A}$ μ-c.t.p., pero T es invertible con inversa medible y $\bigvee_{j=-\infty}^{\infty} T^{-j}\mathcal{P} = \mathcal{A}$ μ-c.t.p.

Se aclara que cuando T es invertible, tanto son particiones generadoras las que cumplen i) como las que cumplen ii).

Teorema 4.5 (Kolmogorov-Sinai) *Si $\mathcal{P}$ es una partición generadora con $H(\mathcal{P}) < \infty$, entonces*

$$h_\mu(T) = h(T, \mathcal{P})$$

Prueba: Consideraremos los dos casos i) y ii) de la defínción de partición generadora

Caso i): Sea $\mathcal{P}_n = \wedge_{j=0}^{n-1} T^{-j}\mathcal{P}$. Tenemos una sucesión de particiones cada vez mas finas, y que generan la σ-álgebra, porque $\mathcal{P}$ es partición generadora.

Aplicando la proposición 3.4 y el lema 2.4, para toda partición finita $\mathcal{Q}$ se cumple:

$$h(T, \mathcal{Q}) - h(T, \mathcal{P}_n) \leq H(\mathcal{Q}|\mathcal{P}_n) \to 0$$

Por otro lado, usando la proposición 3.5 se tiene $h(T, \mathcal{P}_n) = h(T, \mathcal{P})$ para todo $n \geq 1$. Luego, para toda partición finita $\mathcal{Q}$ se tiene

$$h(T, \mathcal{Q}) \leq h(T, \mathcal{P})$$

Finalmente

$$h_\mu(T) = \sup\{h(T,\mathcal{Q}) : \mathcal{Q} \text{ partición finita }\} \leq h(T,\mathcal{P}) \leq \sup\{h(T,\mathcal{R}) : H(\mathcal{R}) < \infty\} = h_\mu(T)$$

Caso ii): Sea $\mathcal{P}_n = \wedge_{j=-n}^{n} T^{-j}\mathcal{P} = T^n\left(\wedge_{j=0}^{2n} T^{-j}\mathcal{P}\right)$. Usando la proposición 3.5 se tiene $h(T,\mathcal{P}_n) = h(T,\mathcal{P})$ para todo $n \geq 1$.

La prueba continúa como en el caso anterior, sin mas cambios. ■

Ejemplos: Entropía métrica del shift de Bernoulli.

Sea $\sigma : B(m) \mapsto B(m)$ el shift unilateral de m símbolos.

Tomemos la partición formada por los m cilindros de radio 1:

$$\mathcal{P} = \{C(1), C(2), \dots, C(m)\}$$

Obsérvese que para todo $n \geq 1$ la partición $\mathcal{P}_n = \wedge_{j=0}^{n-1} T^{-j}\mathcal{P}$ es la de los cilindros de radio n. Como la σ-álgebra de Borel es la generada por los cilindros, la partición $\mathcal{P}$ es generadora.

Luego, *para cualquier medida μ que sea invariante por el shift, la entropía métrica $h_\mu(T)$ es igual a $h(T,\mathcal{P})$.*

Si la medida μ es de Bernoulli, con vector de probabilidad $p = (p(1), p(2), \dots, p(m))$, entonces

$$H(\mathcal{P}_n) = -\sum_{i_1=1}^{m}\sum_{i_2=1}^{m}\dots\sum_{i_n=1}^{m} \mu(C(i_1,i_2,\dots,i_n)) \log \mu(C(i_1,i_2,\dots,i_n))$$

$$= -\sum_{i_1=1}^{m}\sum_{i_2=1}^{m}\dots\sum_{i_n=1}^{m} p(i_1)p(i_2)\dots p(i_n) \log p(i_1)p(i_2)\dots p(i_n) =$$

$$= -\sum_{j=1}^{n}\sum_{i_1=1}^{m}\dots\sum_{i_j=1}^{m}\dots\sum_{i_n=1}^{m} p(i_1)\dots p(i_j)\dots p(i_n) \log p(i_j)$$

Como $\sum_{i=1}^{m} p(i) = 1$, se tiene:

$$H(\mathcal{P}_n) = -\sum_{j=1}^{n}\sum_{i_j=1}^{m} p(i_j)\log p(i_j) = -n\sum_{i=1}^{m} p(i)\log p(i)$$

Dividiendo entre n y haciendo n tender a infinito, se tiene

$$h_\mu(T) = h(T,\mathcal{P}) = -\sum_{i=1}^{m} p(i)\log p(i)$$

Veamos que, entre las medidas que son de Bernoulli, la que da mayor entropía métrica es la que tiene vector de probabilidad $p = (1/m, 1/m, \dots, 1/m)$, y que el shift con esa medida de Bernoulli equidistribuida tiene entropía métrica igual a $\log m$.

En efecto, para el vector de probabilidad $(1/m, 1/m, \dots, 1/m)$, la entropía métrica es

$$-m\frac{1}{m}\log\frac{1}{m} = \log m$$

Para cualquier otro vector de probabilidad, usando la función $\phi(x) = -x\log x$, y la desigualdad de Jensen, se tiene

$$h_\mu(T) = \sum_{i=1}^{m}\phi(p_i) = m\sum_{i=1}^{m}\phi(p_i)/m \leq m\phi\left(\sum_{i=1}^{m}p_i/m\right) = m\phi(1/m) = \log m$$

Ejercicio:

Probar que la entropía métrica del shift de Bernoulli bilateral es $-\sum_{i=1}^{m} p_i \log p_i$

Ejercicio:

Probar que la emtropía métrica del shift de Markov de vector $p = (p_1, p_2, \dots, p_n)$ y matriz $P = (P_{ij})$ es $-\sum_{i=1}^{m}\sum_{j=1}^{m} p_i P_{ij}\log P_{ij}$.

Ejemplo: La entropía métrica de la transformación $T(z) = z^k$, con k natural mayor o igual que 2, en el círculo $S^1 = \{z \in \mathbb{C} : |z| = 1\}$, para la medida de Lebesgue, es $\log k$. En efecto, vimos que T es equivalente por isomorfismo de espacios de medida al shift de Bernoulli con k símbolos, y con vector de probabilidad $(1/k, \dots, 1/k)$.

Ejemplo: La herradura de Smale tiene entropía métrica, respecto a la medida de Bernoulli de vector de probabilidad (1/2, 1/2), igual a $\log 2$. En efecto, es conjugada al shift bilateral de 2 símbolos.

5 Entropía topológica.

En lo que sigue, $T : X \mapsto X$ será una transformación continua en un espacio métrico compacto X.

Definición 5.1 Sean n natural ≥ 1 y $\epsilon > 0$ real, dados.

Dos puntos x e y de X *se* (ϵ, n)*- acompañan* si $\mathrm{d}(T^j(x), T^j(y)) < \epsilon$ para todo $j = 0, 1, \dots, n-1$.

Dos puntos x e y de X *se* (ϵ, n)*-separan* si no se (ϵ, n)-acompañan. Es decir, existe algún $j = 0, 1, \dots, n-1$ tal que $\mathrm{d}(T^j(x), T^j(y)) \geq \epsilon$.

Un subconjunto E de X es *un* (ϵ, n)- *generador* si dado $x \in X$ existe algún $y \in E$ tal que x e y se (ϵ, n)-acompañan.

Un subconjunto no vacío S de X es *un* (ϵ, n)-*separador* si dos puntos x e y de S, diferentes, se (ϵ, n)-separan

Usaremos la siguiente notación: $\sharp A$ = cantidad de elementos del conjunto A $\in [0, \infty]$.

Proposición 5.2 *Dado $\epsilon > 0$ existe $M = M(\epsilon) \geq 1$ tal que para todo $n \geq 1$ existe un (ϵ, n)-generador E con $\sharp E \leq M^n$.*

Prueba: Sea $\mathcal{V}$ un cubrimiento finito de X con bolas abiertas de radio $\epsilon/2$, formado por, digamos, M bolas.

Dada una n-upla $(V_0, V_1, \ldots, V_{n-1})$ de bolas de $\mathcal{V}$, elijamos, cuando no es vacío, un y solo un punto en el conjunto $\cap_{j=0}^{n-1} T^{-j} V_j$. Sea E el conjunto de los puntos así elegidos. Se tiene $\sharp E \leq M^n$.

Veamos que E es un (ϵ, n)-generador. En efecto, dado $x \in X$, existe para cada $j = 0, 1, \ldots, n-1$, algún V_j del cubrimiento $\mathcal{V}$ que contiene al punto $T^j(x)$. Luego $x \in \cap_{j=0}^{n-1} T^{-j} V_j \neq \emptyset$, y existe $y \in E$ en el mismo conjunto intersección. Entonces, $\mathrm{d}(T^j(x), T^j(y)) < \mathrm{diam} V_j = \epsilon$ para todo $j = 0, 1, \ldots, n-1$. ■

Definición 5.3 Como consecuencia de la proposición anterior, existe, para todo $\epsilon > 0$ y todo $n \geq 1$, el número natural

$$r(\epsilon, n) = \min\{\sharp E : E \text{ es un } (\epsilon, n)\text{-generador } \} \leq M(\epsilon)^n$$

Además

$$r(\epsilon, n) \geq 1$$

Consecuencias:

Existe, para todo $\epsilon > 0$, el número real

$$\limsup_{n \to \infty} \frac{\log r(\epsilon, n)}{n} \geq 0$$

Esta expresión decrece cuando ϵ crece. En efecto, si $\epsilon' < \epsilon$, todo (ϵ', n)-generador es un (ϵ, n)-generador. Entonces $r(\epsilon', n) \geq r(\epsilon, n)$. Calculando el logaritmo, dividiendo entre n, y tomando el límite superior cuando n tiende a infinito, se mantiene la desigualdad.

Luego,

$$\lim_{\epsilon \to 0} \limsup_{n \to \infty} \frac{\log r(\epsilon, n)}{n} = \sup_{\epsilon > 0} \limsup_{n \to \infty} \frac{\log r(\epsilon, n)}{n} \in [0, \infty]$$

Proposición 5.4 *Si S es un (ϵ, n)-separador y E es un $(\epsilon/2, n)$-generador, entonces $\sharp S \leq \sharp E$.*

Prueba: Veamos que puede establecerse una correspondencia inyectiva que va de S a E.

Dado $x \in S$, como E es un $(\epsilon/2, n)$-generador, existe $y \in E$ tal que x e y se $(\epsilon/2, n)$-acompañan.

La correspondencia así establecida de S a E es inyectiva, pues si no lo fuera, existirían $x \neq z$ en S que son $(\epsilon/2, n)$- aompañados por el mismo $x \in E$. Entonces, por la propiedad triangular de la distancia, los puntos x y z se (ϵ, n) acompañan entre sí. Esto contradice la hipótesis de que S es un (ϵ, n)- separador. ■

Definición 5.5 Como consecuencia de la proposición anterior, existe para todo $\epsilon > 0$ y todo $n \geq 1$ el número natural

$$s(\epsilon, n) = \max\{\sharp S : S \text{ es un } (\epsilon, n)\text{-separador } \} \leq r(\epsilon/2, n) \leq M(\epsilon/2)^n$$

Además

$$s(\epsilon, n) \geq 1$$

Consecuencias:

Existe, para todo $\epsilon > 0$, el número real

$$\limsup_{n \to \infty} \frac{\log s(\epsilon, n)}{n} \geq 0$$

Esta expresión decrece cuando ϵ crece. En efecto, si $\epsilon > \epsilon'$, todo (ϵ, n)-separador es un (ϵ', n)-separador. Entonces $s(\epsilon, n) \leq s(\epsilon', n)$. Calculando el logaritmo, dividiendo entre n, y tomando el límite superior cuando n tiende a infinito, se mantiene la desigualdad.

Luego,

$$\lim_{\epsilon \to 0} \limsup_{n \to \infty} \frac{\log s(\epsilon, n)}{n} = \sup_{\epsilon > 0} \limsup_{n \to \infty} \frac{\log s(\epsilon, n)}{n} \in [0, \infty]$$

Proposición 5.6 *Si S es un (ϵ, n)-separador con la cantidad máxima de elementos, entonces también es un (ϵ, n)-generador.*

Prueba: Por absurdo, supongamos que existe $x \in X$ tal que para todo $y \in S$, los puntos x e y no se (ϵ, n)-acompañan. Entonces agregando el punto x al conjunto S se tiene otro (ϵ, n)-separador, con más elementos que S. ■

Consecuencias:

$$r(\epsilon, n) \leq s(\epsilon, n) \leq r(\epsilon/2, n)$$

$$\lim_{\epsilon \to 0} \limsup_{n \to \infty} \frac{\log s(\epsilon, n)}{n} = \lim_{\epsilon \to 0} \limsup_{n \to \infty} \frac{\log r(\epsilon, n)}{n} \in [0, \infty]$$

Definición 5.7 Se llama *entropía topológica de la transformación T* a

$$h(T) = \lim_{\epsilon \to 0} \limsup_{n \to \infty} \frac{\log s(\epsilon, n)}{n} = \sup_{\epsilon > 0} \limsup_{n \to \infty} \frac{\log s(\epsilon, n)}{n}$$

$$= \lim_{\epsilon \to 0} \limsup_{n \to \infty} \frac{\log r(\epsilon, n)}{n} = \sup_{\epsilon > 0} \limsup_{n \to \infty} \frac{\log r(\epsilon, n)}{n} \in [0, \infty]$$

El siguiente teorema, cuya demostración no incluiremos, pero puede encontrarse en R. Mañé: Teoría Ergódica, vincula la entropía métrica con la entropía topológica:

Teorema 5.8 (Principio variacional de la entropía) *La entropía topológica de la transformación T continua, del espacio métrico compacto X, es el supremo de las entropías métricas de T para las medidas de probabilidad en los borelianos de X que son T invariantes.*

$$h(T) = \sup\{h_\mu(T) : \mu \in \mathcal{M}_T(X)\}$$

Interpretación de la entropía topológica

Por un lado, $h(T) > 0$ si y solo si para alguna medida μ que sea T-invariante, la entropía métrica $h_\mu(T) > 0$. Ya dimos un significado heurístico a la entropía métrica, como el grado de desorden que provoca T en el espacio.

Demos ahora otra interpretación: Decimos que estudiamos la dinámica con error $\epsilon > 0$ si no distinguimos puntos del espacio que distan menos que ϵ. Así, si x e y se (ϵ, n) acompañan, no distinguiremos la órbita de x de la de y, por lo menos hasta tiempo n.

Luego $s(\epsilon, n)$ es el número máximo de órbitas distinguibles hasta tiempo n con error ϵ. Es el número de datos necesarios para describir la dinámica del sistema hasta tiempo n con error ϵ.

La entropía topológica positiva, quiere decir que el número de datos necesarios para describir la dinámica con error pequeño, crece exponencialmente con el tiempo n.

Veamos que la entropía topológica es invariante por conjugaciones. Es entonces un invariante topológico.

Definición 5.9 Sean $T_1 : X_1 \mapsto X_1$ y $T_2 : X_2 \mapsto X_2$ transformaciones continuas en espacios métricos compactos.

Se dice que T_1 *es semiconjugada con* T_2 si existe $f : X_1 \mapsto X_2$ continua y sobreyectiva que cumple $f \circ T_1 = T_2 \circ f$.

Se dice que T_1 *es conjugada con* T_2 si existe $f : X_1 \mapsto X_2$ homeomorfismo que cumple $f \circ T_1 = T_2 \circ f$.

Se observa que cuando dos transformaciones son conjugadas, cada una de ellas es semiconjugada con la otra.

Nótese que $f \circ T_1 = T_2 \circ f$ implica $f \circ T_1^j = T_2^j \circ f$, para todo $j \geq 0$.

Teorema 5.10 *Si T_1 es semiconjugada con T_2 entonces $h(T_1) \geq h(T_2)$.*

Si T_1 es conjugada con T_2 entonces $h(T_1) = h(T_2)$.

Prueba: La primera afirmación implica la segunda.

Sea $f : X_1 \mapsto X_2$ continua y sobreyectiva, según la definición de semiconjugación.

Como f es sobreyectiva, existe $g : X_2 \mapsto X_1$, inversa a la derecha de f. Luego g es inyectiva.

Como f es continua, dado $\epsilon > 0$, existe $\delta > 0$ tal que $\mathrm{d}(x, y) < \delta$ en X_1 implica $\mathrm{d}(f(x), f(y)) < \epsilon$ en X_2.

Como $f \circ T_1^j = T_2^j \circ f$, para todo $j \geq 0$, todo (ϵ, n)-separador en X_2 para T_2, es transformado por g en un (δ, n)-separador en X_1 para T_1. Luego:

$$s_2(\epsilon, n) = \max\{\sharp S : S \text{ es un } (\epsilon, n)\text{-separador en } X_2 \text{ para } T_2\}$$

$$\leq \max\{\sharp S : S \text{ es un } (\delta, n)\text{-separador en } X_1 \text{ para } T_1\} = s_1(\delta, n)$$

Entonces, para todo $\epsilon > 0$:

$$\limsup_{n\to\infty} \frac{\log s_2(\epsilon, n)}{n} \leq \limsup_{n\to\infty} \frac{\log s_1(\delta, n)}{n} \leq \sup_{\rho>0} \limsup_{n\to\infty} \frac{\log s_1(\rho, n)}{n} = h(T_1)$$

Haciendo ϵ tender a cero:

$$h(T_2) \leq h(T_1)$$

■

Proposición 5.11 *Si $A \subset X$ es compacto e invariante para adelante (esto es $T(A) \subset A$), entonces*

$$h(T) \geq h(T|_A)$$

Prueba: Todo (ϵ, n) separador para $T|_A$ es un (ϵ, n) separador para T. Entonces

$$s_A(\epsilon, n) = \max\{\sharp S : S \text{ es un } (\epsilon, n)\text{-separador para } T|_A\}$$

$$\leq \max\{\sharp S : S \text{ es un } (\epsilon, n)\text{-separador para } T\} = s(\epsilon, n)$$

Por definición de entropía topológica, lo anterior implica la tesis. ■

Veamos una condición muy utilizada para demostrar que una transformación tiene entropía topológica positiva:

Definición 5.12 Un transformación $T : X \mapsto X$ *contiene una herradura* si existe un compacto A invariante hacia adelante tal que $T|_A$ es semiconjugada con la herradura de Smale.

Corolario 5.13 *Si una transformación contiene una herradura, entonces tiene entropía topológica mayor o igual que* $\log 2$*.*

Prueba: Para la medida de Bernoulli con vector de probabilidad $(1/2, 1/2)$, la herradura de Smale σ tiene entropía métrica igual a $\log 2$. Por el principio variacional de la entropía, la entropía topológica de la herradura de Smale es mayor o igual que $\log 2$.(Mas adelante veremos que es igual a $\log 2$).

Luego

$$h(T) \geq h(T|_A) \geq h(\sigma) \geq \log 2$$

■

El siguiente teorema muestra como la entropía topológica crece al iterar T:

Teorema 5.14 *Sea T continua en un espacio métrico compacto. Se cumple:*

1. *$h(T^m) = mh(T)$ para todo $m \geq 1$.*
2. *Si T es un homeomorfismo, entonces $h(T^{-1}) = h(T)$ y $h(T^{-m}) = mh(T)$ para todo $m \geq 1$.*

Prueba:

1) Todo (ϵ, n)-separador de T^m es un (ϵ, mn)-separador de T. Entonces:

$$s_m(\epsilon, n) = \max\{\sharp S : S \text{ es un } (\epsilon, n)\text{-separador para } T^m\}$$

$$\leq \max\{\sharp S : S \text{ es un } (\epsilon, mn)\text{-separador para } T\} = s(\epsilon, mn)$$

Luego:

$$\limsup_{n\to\infty} \frac{\log s_m(\epsilon, n)}{n} \leq \limsup_{n\to\infty} \frac{m \log s(\epsilon, mn)}{mn} \leq m \limsup_{k\to\infty} \frac{\log s(\epsilon, k)}{k}$$

Tomando $\epsilon \to 0$, se obtiene:

$$h(T^m) \leq mh(T)$$

Para probar la desigualdad contraria, usemos la definición de entropía topológica basada en los generadores.

Dado $\epsilon > 0$, por la continuidad de T, sea $\delta > 0$ tal que, si $\mathrm{d}(x, y) < \delta$, entonces, para $j = 0, 1, \ldots, m-1$, se cumple $\mathrm{d}(T^j(x), T^j(y)) < \epsilon$.

Todo (δ, n)-generador para T^m será un $(\epsilon, mn+j)$-generador para T. Luego

$$r_m(\delta, n) = \min\{\sharp E : E \text{ es un } (\delta, n)\text{-generador para } T^m\}$$

$$\geq \min\{\sharp E : E \text{ es un } (\epsilon, mn+j)\text{-generador para } T\} = r(\epsilon, mn+j)$$

$$r_m(\delta, n) \geq \max_{0 \leq j \leq m-1} r(\epsilon, mn+j)$$

Luego, para todo $\epsilon > 0$, se cumple:

$$h(T^m) = \sup_{\rho > 0} \limsup_{n \to \infty} \frac{\log r_m(\rho, n)}{n} \geq \limsup_{n \to \infty} \frac{\log r_m(\delta, n)}{n}$$

$$\geq \limsup_{n \to \infty} \frac{\log\left(\max\{r(\epsilon, mn+j) : 0 \leq j \leq m-1\}\right)}{n}$$

$$\geq \limsup_{N \to \infty} \frac{\log r(\epsilon, N)}{\text{ent } (N/m)} = m \limsup_{N \to \infty} \frac{\log r(\epsilon, N)}{N}$$

Haciendo $\epsilon \to 0$, se tiene

$$h(T^m) \geq m h(T)$$

2) Todo (ϵ, n)-generador de T es transformado por T^{n-1} en un (ϵ, n)- generador de T^{-1}, con la misma cantidad de elementos. Y recíprocamente, todo (ϵ, n)-generador de T^{-1}, es transformado por $T^{-(n-1)}$ en un (ϵ, n)-generador de T.

Luego la cantidad mínima de elementos de los (ϵ, n)-generadores de T y de T^{-1} son iguales. Por definición de entropía topológica, $h(T^{-1}) = h(T)$.

La última igualdad es una consecuencia inmediata de las anteriores. ■

6 Ejemplos de cálculo de la entropía topológica

1: **La entropía topológica del shift de m símbolos es $\log m$.**

Demostrémoslo para el shift unilateral σ. Se deja como ejercicio, adaptar la prueba, para verificar que lo mismo vale para el shift bilateral.

1a. parte) $h(\sigma) \geq \log m$. Usando la entropía métrica del shift de Bernoulli y el principio variacional de la entropía, se deduce esta desigualdad. Pero, ya que no demostramos este principio, probemos la desigualdad directamente a partir de la definición de entropía topológica.

Elijamos para cada $n \geq 1$ un conjunto S formado por un y solo un punto en cada cilindro $C(a_0, a_1, \dots, a_{n-1})$. Se tiene $\sharp S = m^n$, y S es un (ϵ, n)-separador, para todo $\epsilon > 0, \epsilon < 1$. En efecto: $x \neq y \in S$ implica que existe $j = 0, 1, \dots, n-1$ tal que $x_j \neq y_j$. Luego:

$$\mathrm{d}(\sigma^j x, \sigma^j y) = \sum_{i=1}^{\infty} \frac{|x_{i+j} - y_{i+j}|}{\lambda^i} \geq |x_j - y_j| \geq 1 > \epsilon$$

Entonces $s(\epsilon, n) \geq m^n$ para todo $\epsilon < 1$, de donde $h(\sigma) \geq \log m$.

2a. parte) $h(\sigma) \leq \log m$.

Dado $\epsilon > 0$ sea N tal que todos los cilindros de radio N tienen diámetro menor que ϵ.

Elijamos para cada $n > N$ un conjunto E formado por un y solo un punto en cada cilindro $C(a_0, a_1, \dots, a_{n-1})$. Se tiene $\sharp E = m^n$, y E es un $(\epsilon, n-N)$-generador, para todo $\epsilon > 0$. En efecto: dado $x \in B(m)$, existe $y \in E$ en el cilindro $C(x_0, x_1, \dots, x_{n-1})$. Luego, para todo $j = 0, 1, \dots, n-N$, los puntos $\sigma^j x$ y $\sigma^j y$ están en el cilindro $C(x_j, \dots, x_{n-1})$ de radio $n - j \geq N$. Entonces están en el mismo cilindro de radio N, y distan menos que ϵ. Los puntos x e y se $(\epsilon, n-N)$-acompañan.

Concluimos que $r(\epsilon, n-N) \leq m^n$, de donde

$$\limsup_{n \to \infty} \frac{\log r(\epsilon, n-N)}{n-N} \leq \lim_{n \to \infty} \frac{n}{n-N} \log m = \log m$$

Luego $h(\sigma) \leq \log m$.

2: **La entropía topológica de la herradura de Smale (con 2 patas) es $\log 2$.**

La herradura de Smale (con dos patas) es conjugada al shift bilateral de dos símbolos, que tiene entropía topológica $\log 2$. La entropía topológica es invariante por conjugaciones.

3: **La entropía topológica de la transformación $T(z) = z^k$ en el círculo S^1, es $\log k$.**

La entropía métrica para la medida de Lebesgue es $\log k$, porque T es equivalente, por isomorfismo de espacios de medida, al shift de Bernoulli con vector de probabilidad $(1/k, \dots, 1/k)$. Luego $h(T) \geq \log k$, por el principio variacional de la entropía.

Pero veámoslo directamente, usando separadores: Dado $n \geq 1$, sea S el conjunto de puntos extremos de $2k^n$ intervalos iguales en S^1, de longitud $1/2k^n$ cada uno. Se tiene $\sharp S = 2k^n$, y S es un (ϵ, n) separador para todo $\epsilon < 1/2k$. En efecto, dos puntos $x \neq y \in S$, distan por lo menos $1/2k^n$. Aplicando T, la distancia, mientras sea menor que 1/2k, se multiplica por k. Luego, para algún $j = 0, 1, \dots, n-2$ se cumple $\mathrm{d}(T^j x, T^j y) \geq 1/2k$, o de lo contrario

$$\mathrm{d}(T^{n-1}x, T^{n-1}y) = k^{n-1}\ \mathrm{d}(x, y) \geq \frac{k^{n-1}}{2k^n} = \frac{1}{2k} > \epsilon$$

Luego $s(\epsilon, n) \geq 2k^n$, y aplicando la definición de entropía topológica, se tiene $h(T) \geq \log k$.

Ahora demostremos que $h(T) \leq \log k$. Dado $\epsilon > 0$, dividamos S^1 en $M = M(\epsilon)$ intervalos iguales de longitud $1/M < \epsilon$. Dado $n \geq 1$, dividamos cada intervalo en k^n intervalitos iguales de longitud $< \epsilon/k^n$. Sea E el conjunto de puntos extremos de esos intervalitos. Se tiene $\sharp E = Mk^n$. Afirmamos que E es un (ϵ, n)-generador. En efecto, dado $x \in S^1$, sea $y \in E$ el punto mas próximo a x. Para todo $j = 0, 1, \ldots, n-1$ se cumple:

$$\mathrm{d}(T^j x, T^j y) \leq k^j \ \mathrm{d}(x, y) \leq \frac{k^j \epsilon}{k^n} < \epsilon$$

Luego, $r(\epsilon, n) \leq Mk^n$. Aplicando la definición de entropía topológica, se deduce $h(T) \leq \log k$.

4: **La entropía topológica de cualquier traslación en el toro es nula.**

La traslación T en el toro conserva la distancia. Dado ϵ sea $\mathcal{V}$ un cubrimiento del toro con $M = M(\epsilon)$ bolas de radio ϵ. Sea E el conjunto de centros de las bolas de $\mathcal{V}$. Se tiene $\sharp E = M$. Además E es un (ϵ, n)- generador, para todo $n \geq 1$. En efecto, dado x en el toro, existe alguna bola de $\mathcal{V}$ que contiene a x. Luego, existe algún $y \in E$, tal que $\mathrm{d}(x, y) < \epsilon$. Pero como al aplicar T no se modifica la distancia, se tiene $\mathrm{d}(T^j x, T^j y) < \epsilon$ para todo $j \geq 0$.

De lo anterior se deduce que $1 \leq r(\epsilon, n) \leq M$ independiente de n. Luego, tomando logaritmo, y dividiendo entre n, su límite es cero, para todo $\epsilon > 0$. Luego $h(T) = 0$.

Nótese que lo anterior demuestra que la entropía topológica de cualquier isometría en un espacio métrico compacto es nula.

5: **La entropía topológica de cualquier homeomorfismo T del círculo S^1 es nula.**

Consideraremos el círculo S^1 como el cociente $\mathbb{R}/\sim$, con $x \sim y$ en $\mathbb{R}$ si $y - x$ es entero. Orientaremos el círculo, según valores crecientes de x en $\mathbb{R}$.

Basta demostrar la tesis asumiendo que T preserva la orientación del círculo. En efecto, como $h(T^2) = 2h(T)$, si T no preserva la orientación de S^1, podemos sustituirlo por T^2 que preserva la orientación.

Sea dado ϵ positivo, $\epsilon < 1/2$.

Sea F el conjunto de puntos extremos de $M = M(\epsilon)$ intervalos iguales en S^1, de longitud $1/M < \epsilon$ cada uno. Tenemos que F es un $(\epsilon, 1)$ generador tal que, dado $x \in S^1$, existe $y \geq x, y \in F$ tal que $y - x < \epsilon$. Demostremos que para todo $n \geq 1$ existe un (ϵ, n)-generador con a lo sumo nM elementos, y tal que dado $x \in S^1$, existe $y \geq x, y \in E$ tal que para todo

$j = 0, 1, \ldots, n-1$ se cumple $0 \leq T^j y - T^j x < \epsilon$. La demostración se hará por inducción en n:

Sea E un (ϵ, n) -generador que cumple lo anterior.

Afirmamos que $E \cup T^n F$ es un $(\epsilon, n+1)$-generador que también cumple lo anterior. En efecto, dado $x \in S^1$, sea $y \geq x, y \in E$ tal que, para $j = 0, 1, \ldots, n-1$ se cumple $T^j y - T^j x < \epsilon$. Ahora, hay dos posibilidades:

1) o bien $T^n y - T^n x < \epsilon$, en cuyo caso, x e y se $(\epsilon, n+1)$-acompañan,

2) o bien $T^n y - T^n x \geq \epsilon$. En este caso, existe $z \in F$ tal que $T^n x \leq z < T^n y$ y $z - T^n x < \epsilon$. Como T preserva la orientación, para todo $i = 0, 1, \ldots, n$ se cumple $T^{n-i} x \leq T^{-i} z < T^{n-i} y$; luego $T^{-i} z - T^{n-i} x < T^{n-1} y - T^{n-i} x < \epsilon$. Entonces, el punto $T^n z \in T^n F$ $(\epsilon, n+1)$-acompaña a x.

Como $\sharp(E \cup T^n F) \leq \sharp E + M \leq (n+1)M$, se tiene demostrada la afirmación.

Entonces $r(\epsilon, n) \leq nM$. Tomando logaritmo y dividiendo entre n, su límite, cuando $n \to \infty$ es cero. Luego $h(T) = 0$.

6: **La entropía topológica del** $\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$ **en el toro es** $\log\left(\frac{3+\sqrt{5}}{2}\right)$.

El toro T^2 es el cociente $R^2/\sim$ con la relación de equivalencia $x \sim y$ si $x - y \in \mathbb{Z}^2$. Denotamos como $\pi(x)$ o como $\overline{x}$ al punto del toro que es la clase de equivalencia de $x \in \mathbb{R}^2$.

La distancia entre dos puntos $\overline{x}$ y $\overline{y}$ en el toro es igual a la mínima distancia en $\mathbb{R}^2$ de sus representantes x e y. Luego $\mathrm{d}(\overline{x}, \overline{y}) \leq \mathrm{d}(x, y)$, para todos x e y en $\mathbb{R}^2$. Pero además, si $\mathrm{d}(x, y) < 1/2$ entonces $\mathrm{d}(\overline{x}, \overline{y}) = \mathrm{d}(x, y)$.

Llamando A a la matriz $\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}$, se tiene $T(\overline{x}) = \pi(Ax)$.

Sean $\lambda = (3 + \sqrt{5})/2 > 1$ y $\mu = (3 - \sqrt{5}/2) < 1$ los dos valores propios de A. Sean u y v dos vectores propios de A de valores propios respectivos λ y μ. En $\mathbb{R}^2$, la dirección de u, que se dilata con factor $\lambda > 1$, se llama *inestable,* y la dirección de v, que se contrae con factor $\mu < 1$, se llama *estable.*

Para comprender la prueba que sigue, por favor vaya dibujando en $\mathbb{R}^2$ los conjuntos que se definen:

1a. parte: Prueba de $h(T) \geq \log \lambda$.

Dado $\epsilon > 0, \epsilon < 1/2\lambda$, sea I un segmento en $I\!R^2$, según la dirección inestable, con longitud igual a ϵ. Como su longitud es menor que $1/2$, se proyecta inyectivamente en el toro.

Dado $n \geq 1$ se partirá I en s partes iguales, donde $s = \text{ent}(\lambda^n)$. La longitud de cada subintervalo obtenido es $\epsilon/s \geq \epsilon/\lambda^n$.

Sea S el subconjunto formado por los $s+1$ puntos extremos de los subintervalos construidos en I. Se tiene $\sharp S = \text{ent}(\lambda^n) + 1 \geq \lambda^n$.

Afirmamos que $\pi(S)$ es un $(\epsilon, n+1)$-separador en el toro.

En efecto, aplicando A a dos puntos x e y diferentes de S, como son extremos de un segmento con la dirección inestable, se alejarán multiplicando su distancia por $\lambda > 1$. Luego $\text{d}(A^n x, A^n y) = \lambda^n \ \text{d}(x, y) \geq \epsilon$.

Pero esa distancia es en $I\!R^2$ y no en el toro. Para ver que se distancian por lo menos ϵ al aplicar T en el toro, tomemos el primer $J \geq 0$ tal que en el plano $\text{d}(A^J x, A^J y) \geq \epsilon$. Se tiene $0 \leq J \leq n$.

Si $J = 0$, entonces $\text{d}(x, y) = \epsilon < 1/2$, porque la longitud de I es ϵ. Entonces, en el toro $\text{d}(\overline{x}, \overline{y}) = \text{d}(x, y) = \epsilon$.

Si $J \geq 1$, entonces $\text{d}(A^{J-1}x, A^{J-1}y) < \epsilon < 1/2\lambda$. Al aplicar A una vez más, la distancia se multiplica por λ, luego $\text{d}(A^J x, A^J y) < 1/2$, y se tiene:

$$\text{d}(T^J \overline{x}, T^J \overline{y}) = \text{d}(A^J x, A^J y) \geq \epsilon$$

Esto prueba que $\pi(S)$ es un $(\epsilon, n+1)$ separador en el toro.

Entonces $s(\epsilon, n+1) \geq \sharp\pi(S) = \sharp S \geq \lambda^n$. Tomando logaritmo, dividiendo entre n y haciendo tender n a infinito, se tiene $h(T) \geq \log \lambda$.

2a. parte: Prueba de $h(T) \leq \log \lambda$.

Sea dado $\epsilon > 0, \epsilon < 1/2$. Sea en $I\!R^2$ un *cuadriculado* formado por infinitas rectas paralelas a la dirección inestable, e infinitas rectas paralelas a la dirección estable, tales que los lados de los rombitos que forman sean exactamente $\epsilon/2$.

Sea Q el conjunto (finito) de los vértices del cuadriculado que están en el rectángulo $[-2, 3]^2$. Sea $q = \sharp Q$.

Dado $n \geq 1$ dividamos cada lado inestable de los rombitos en r segmentos iguales con $r = \text{ent}(\lambda^n) + 1$. Los lados estables no los dividimos. Cada subintervalo inestable así obtenido tiene longitud $\epsilon/2r < \epsilon/2\lambda^n$.

Sea E el conjunto finito de los puntos extremos de los intervalos así construidos, que están en el rectángulo $[-1,2]^2$. Se tiene $\sharp E \leq rq \leq (\lambda^n + 1)q$.

El conjunto $\pi(E)$ en el toro tiene a lo sumo $(\lambda^n + 1)q$ puntos, con q independiente de n.

Afirmamos que $\pi(E)$ es un (ϵ, n)-generador. En efecto, dado $\overline{x} \in T^2$, sea $x \in I\!R^2$ su único representante en $[0,1)^2$. Sea $y \in E$ el punto mas próximo de x. Tracemos la recta estable por x y la inestable por y, que se cortan en z. Por construcción, el segmento $[x,z]$ es estable y su longitud es menor que $\epsilon/2$. Análogamente el segmento $[z,y]$ es inestable y su longitud es menor que $\epsilon/2\lambda^n$. Aplicando la matriz A, los segmentos inestables se multiplican por $\lambda > 1$ y los estables por $\mu < 1$, de donde, en $I\!R^2$ se obtiene:

$$\mathrm{d}(A^j x, A^j y) \leq \mathrm{d}(A^j x, A^j z) + \mathrm{d}(A^j z, A^j y) = \mu^j \, \mathrm{d}(x,z) + \lambda^j \, \mathrm{d}(z,y)$$

$$\leq \mathrm{d}(x,z) + \frac{\lambda^j}{\lambda^n}\frac{\epsilon}{2} < \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon$$

si $j = 0, 1, \ldots, n-1$. Como $\mathrm{d}(T^j\overline{x}, T^j\overline{y}) \leq \mathrm{d}(A^j x, A^j y)$, el conjunto $\pi(E)$ es un (ϵ, n)-generador.

Entonces $r(\epsilon, n) \leq (\lambda^n + 1)q$. Tomando logaritmo, dividiendo entre n y haciendo $n \to \infty$ se obtiene $h(T) \leq \log \lambda$.